FEBS Letters 582 (2008) 3515–3519

Nuclear matrix contains novel WD-repeat and disordered-region-rich proteins

Kohei Ishii[a], Yasuhiro Hirano[b], Nanase Araki[a], Toshiyuki Oda[b], Masahiro Kumeta[b], Kunio Takeyasu[b], Kazuhiro Furukawa[a], Tsuneyoshi Horigome[a,*]

[a] *Graduate School of Science and Technology, Niigata University, Igarashi-2, Nishi-ku, Niigata 950-2181, Japan*
[b] *Graduate School of Biostudies, Kyoto University, Yoshida-tachibanacho, Kyoto 606-8303, Japan*

Received 15 May 2008; revised 8 August 2008; accepted 10 September 2008

Available online 18 September 2008

Edited by Ulrike Kutay

Abstract To find novel proteins predicted to participate in the formation of nuclear bodies, nuclear speckles, and nuclear macro-protein complexes, we applied proteome analysis to a HeLa cell nuclear matrix fraction. Proteins in the fraction were separated by SDS–PAGE, digested with trypsin, and analyzed by nanoflow liquid chromatography–iontrap-tandem mass spectrometry. Three hundred and thirty three proteins including 39 novel ones were identified. Seven WD-repeat proteins and 16 disordered region-rich proteins, which act frequently as scaffolding proteins for macro-protein complexes, were found amongst the novel proteins.

Keywords: Disordered region; Nuclear architecture; Nuclear matrix; Nuclear proteome; WD-repeat

1. Introduction

The inner nuclear structure of the cell is dynamic and contains various distinct domains that are closely related to the nuclear function. For the establishment of specific domains within the nucleus, high-order organization of the nuclear interior is necessary. Proteins that participate in particular nuclear functions are not equally distributed throughout the whole nuclear interior, but are rather localized to specific compartments as nuclear bodies and speckles [1–3].

The complex organization of the inner nuclear structure has been explained by the concept of a structural framework of a karyoskeleton or a matrix that might contribute by spatially organizing the nuclear compartments [4]. However, the molecular identity of such a structural scaffold remains elusive. The self-organization model of nuclear assembly suggests that nuclear bodies are largely the result of the sum of many, likely transient and non-specific, interactions amongst the resident proteins [5]. The regulated-exchange model also suggests that nuclear speckles are formed through a process of self-assembly, and might not depend on an underlying scaffold structure [2]. However, it has been found that a nuclear body, the RUNX body, comprises a scaffold protein RUNX and many other associated proteins [6]. AML-1, ALL-1, PCNA, and BRCA1 also act as scaffolding proteins in specific nuclear bodies [1]. Similarly, other nuclear bodies may also contain specific scaffolding proteins. Therefore, searches for and analyses of novel nuclear scaffolding proteins and structural proteins are important to gain an insight into the relation between inner nuclear structure and nuclear functions.

In this study, we analyzed a nuclear matrix fraction of HeLa cells by proteome analysis and identified 333 proteins including 39 novel ones. Many candidates for scaffolding proteins were found amongst these proteins.

2. Materials and methods

2.1. Preparation of a HeLa cell nuclear matrix fraction

A ribonucleoprotein-containing nuclear matrix fraction was prepared from HeLa S3 cells according to the method of Fey et al. [7] as shown in Supplementary procedure S1. The specificity of our isolation procedure was demonstrated by Western blotting of different marker proteins, as shown in Supplementary Fig. S1. This nuclear matrix fraction is known to contain the ribonucleoprotein-containing nuclear matrix and cytoplasmic filaments [7]. In this study, we simply call this fraction the "nuclear matrix fraction".

2.2. Separation and identification of proteins in the nuclear matrix fraction

A HeLa cell nuclear matrix fraction (100 μg) was separated by SDS–PAGE. The gel was stained with Coomassie brilliant blue (CBB), and then cut into 29 slices. The slices were treated with trypsin and the resulting tryptic peptides were extracted as described previously [8]. Peptide mixtures were analyzed with an LC–MS/MS (Agilent 1100 LC/MSD Trap XCT Ultra) equipped with an enrichment column and an analytical column (150 × 0.075 mm, ZORBAX 300 SB-C18, 3.5 μm). The mass spectrometer was operated over the range of 350–1800 *m/z*. Protein identification was performed with a Spectrum Mill MS Proteomics Workbench platform (Agilent Technologies, Santa Clara, CA) according to the procedure of Yoshida et al. with small modifications [9]. Peptides matches with a peptide score of more than 6 and a scored peak intensity (SPI) of more than 70% were considered significant. For more details, refer to Supplementary procedure S2.

2.3. Expression of green fluorescence protein (GFP)-fused proteins

cDNA was amplified by PCR against a cDNA library of human testis (Takara, Tokyo). The PCR products were inserted into the pEGFP-C2 vector (Clontech, Mountain View, CA). The cDNAs were introduced into HeLa S3 cells using Effecten Transfection Reagent (Qiagen, Valencia, CA), followed by incubation for 24 h. Cells were washed

*Corresponding author. Fax: +81 25 262 6160.
E-mail address: thori@chem.sc.niigata-u.ac.jp (T. Horigome).

Abbreviations: CBB, Coomassie brilliant blue; GFP, green fluorescence protein; GO, gene ontology; LC–MS/MS, nano-flow liquid chromatography–iontrap-tandem mass spectrometry

0014-5793/$34.00
doi:10.1016/j.febslet.2008.09.019

with PBS, fixed with 4% paraformaldehyde, counter-stained with Hoechst33342, and then observed under a fluorescence microscope (BX50-FLA, Olympus, Tokyo).

3. Results and discussion

3.1. Identification of proteins in the nuclear matrix fraction

A nuclear matrix fraction of HeLa cells was separated by SDS–PAGE and stained with CBB (Fig. 1A). The gel was cut into 29 slices and the proteins in each slice were digested with trypsin. The resulting peptides were extracted from the gel slices and analyzed by LC–MS/MS. Then, 333 proteins were identified in the nuclear matrix fraction (Supplementary Table S1). Well known nuclear matrix proteins: SAF-A SAF-B, fibrillarin, NuMA, matrin, nucleophosmin, nuclear matrix protein p84, many hnRNPs and others, were included in the fraction.

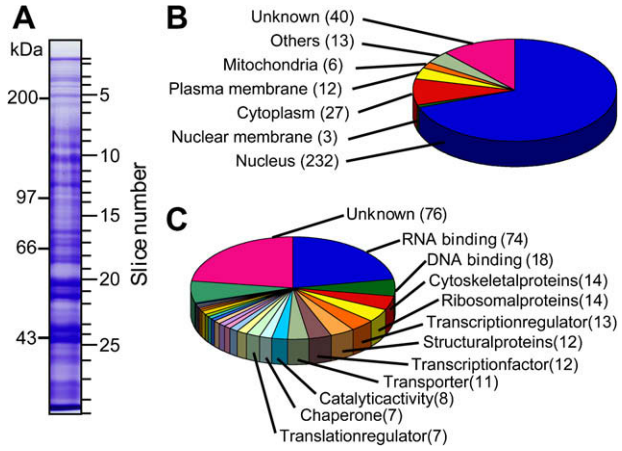

Fig. 1. Subcellular localization and functions of proteins identified in a HeLa cell nuclear matrix fraction. (A) A HeLa cell nuclear matrix fraction was separated by 8% SDS–PAGE, and the gel was stained with CBB and cut into 29 slices. Proteins in the gel slices were digested with trypsin and then identified by LC-MS/MS. (B and C) All 333 proteins identified in the nuclear matrix fraction were classified as to their subcellular localizations (B) and their molecular functions (C) with Quick GO. The numbers in parentheses are the numbers of proteins in the categories.

A literature search was carried out for the 333 proteins with Quick Gene Ontology (GO) based at the EBI (see geneontology.org), as well as the annotation of GO to UniProt and InterPro. About 70% of the identified proteins (235 of 333 proteins) were nuclear ones (Fig. 1B). A half of the proteins classified as "Cytoplasm" ones were cytoskeletal ones (see below). It has been reported that the thus prepared nuclear matrix fraction contains cytoskeletal proteins [7]. Other proteins, of which the major localization sites are the cytoplasm, plasma membrane, mitochondria, etc., are likely to be contaminants arising during preparation of the nuclear matrix fraction. However, a part of these proteins may be localized partially or transiently in the nucleus, as known in other cases [10,11].

Molecular function in GO (Fig. 1C) predicted that 257 of the 333 proteins (77%) possessed at least one known biological function, whereas the biological functions of the remaining 76 proteins (23%) could not be well defined. Among the proteins with certain biological functions, 74 and 18 were classified as RNA-binding and DNA-binding ones, respectively. The RNA-binding protein group contained heterogeneous nuclear ribonucleoproteins and ribosomal proteins. This is reasonable because it is well known that heterogeneous nuclear ribonucleoproteins are abundant in nuclei, form large complexes with pre-mRNA, and play a role in the processing of pre-mRNA and nuclear export of mRNA. The cytoskeletal and structural protein groups contain lamin A, lamin B1, lamin C, vimentin, keratins, actin, clathrin, and others.

For 21 of the 333 proteins, both the localization and function are not shown in GO (Table 1). We randomly selected six of these 21 proteins and then transiently expressed them as GFP-fused proteins in HeLa cells. The fluorescence ap-

Table 1
Uncharacterized proteins found in a HeLa cell nuclear matrix fraction

NCBI accession no.	Gene symbol	Protein name	Disordered region (%)
27469365	C19orf21	Chromosome 19 open reading frame 21	53
62868208	GIYD1	GIY-YIG domain containing 1 isoform 1	19
8749097	HSAPC045	HSPC045 protein	54
55749644	KIAA0701	Hypothetical protein LOC23074 isoform a	27
29826296	MAGEF1	Melanoma antigen family F1	40
12697963	KIAA1709	KIAA1709 protein	20
66773336	C20orf117	KIAA0889 protein	34
55962598	C10orf120	Unnamed protein	54
55958319	FAM27E1	Unnamed protein	80
33341782	PP6569	PP6569	16
Q113424747	AHNAK2	PREDICTED: chromosome 14 open reading frame 78	42
89041269	–	PREDICTED: similar to CDRT15 protein	28
88951501	–	PREDICTED: similar to cytoplasmic beta-actin	30
51467300	–	PREDICTED: similar to FK506-binding protein 4	27
51467395	–	PREDICTED: similar to PRED65	32
24657522	UBQC	Ubiquitin C	41
2001609	LOC647107	Unnamed protein	100
10434310	–	Unnamed protein	7
10433642	–	Unnamed protein	40
55961509	WDR46	WD repeat domain 46	33
31542526	WDSOF1	WD repeats and SOF1 domain containing	35

Proteins of which the functions and subcellular localizations are not shown in GO were selected from Supplementary Table S1 and designated as "uncharacterized proteins". Disordered region (%) was calculated from values predicted with disEMBL.

peared in nucleoli (GFP-KIAA1709 and -WDR46), or nuclei (GFP-GIYD1 and -WDSOF1), or partially in nuclei (GFP-MAGEF1 and -pp6569) (Fig. 2). When, these cells were treated by the procedure for nuclear matrix preparation (Supplementary procedure S1) on cover glasses, GFP-KIAA1709, -WDR46, -GIYD1, -WDSOF1, and -MAGEF1 at least partially remained in the nuclei, and GFP-pp6569 in the cytoplasm (Supplementary Fig. S2), These results suggested that the majority of the uncharacterized proteins are nuclear matrix ones.

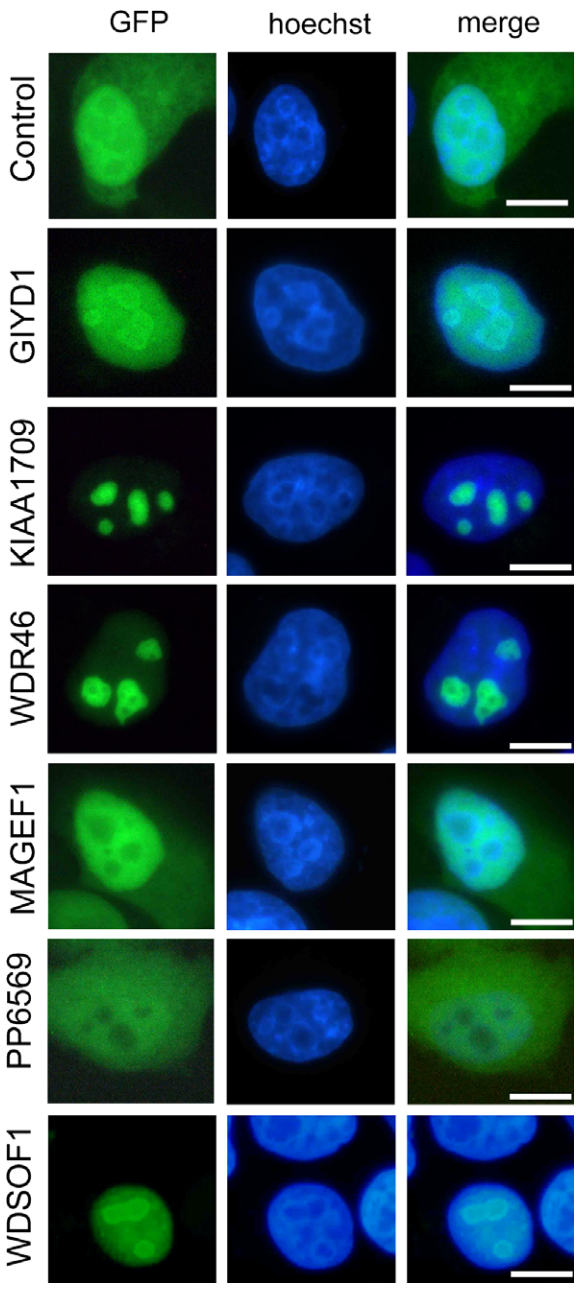

Fig. 2. Localization of novel proteins fused to GFP in HeLa cells. Several novel proteins were expressed as GFP-fused proteins. Cells were fixed, counter-stained with Hoechst and then observed by fluorescence microscopy. Control, GIYD1, WDR46, MAGEF1, and pp6569 indicate GFP, GFP-GIYD1, GFP-WDR46, GFP-MAGEF1, and GFP-pp6569, respectively. Bars: 10 μm.

3.2. WD-repeat proteins

The identified 333 proteins included 21 uncharacterized ones (Table 1), for 18 of which only the localization was predicted by the electric annotation in GO (Table 2). We call these 39 proteins "novel proteins" in this study. The motifs in these proteins were searched for with SMART and 7 were found to contain WD-repeats (Fig. 3). WD-repeats are conserved domains comprising approximately 40–60 amino acids [12] and form a discriminatory conformation termed the β-propeller. Recent whole-genome sequence analysis indicated that there are 260 WD-repeat proteins in man [13]. The presence of 7 WD-repeat proteins amongst the 39 novel proteins (18%), and 13 amongst the 333 nuclear matrix fraction proteins (4%) corresponded to the very high content found when these values were compared with those of the 260 amongst the total 25 077 human proteins (1.0%). These results suggested that the WD-repeat protein group is one of the characteristic groups in the nuclear matrix fraction. WD-repeats act as structural platforms for protein–protein interactions [14,15]. The underlying common function of all WD-repeat proteins is the coordination of multiprotein complex assemblies [16,17]. These repeating units are believed to serve as a scaffold for protein interactions. Furthermore, sof1, an yeast homologue of WDSOF1, is known to be a component of a large ribonucleoprotein complex [18]. Therefore, studies on binding proteins and the functions of the 7 WD-repeat proteins (Fig. 3) will provide a new insight into the structure and function of nuclear bodies, nuclear speckles, and macro-protein complexes.

3.3. Disordered-region-rich proteins

Recent studies revealed that proteins, which are rich in disordered regions, act as assemblers that assemble, stabilize and regulate large multi-protein complexes [19]. A disordered region does not take on any specific conformation, and it can serve as the structural basis for hub proteins and functions as a flexible linker between functional domains [20]. BRCA1, ALL-1 and SAF-B, known as scaffold proteins for multi-protein complexes [21,22], possess high contents of disordered regions (44%, 69%, and 65%, respectively) within their amino acid sequences.

An algorithm, DisEMBL (http://dis.embl.de/), was applied to the 333 nuclear matrix fraction proteins to predict the

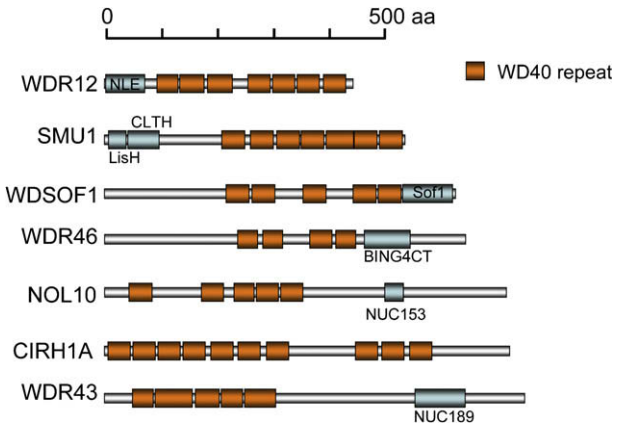

Fig. 3. Novel WD-repeat proteins found in the HeLa cell nuclear matrix fraction. Seven novel proteins containing WD-repeats are shown schematically with other motifs and gene symbols. The scale bar indicates the numbers of amino acid residues.

Table 2
Poorly characterized proteins found in a HeLa cell nuclear matrix fraction

NCBI accession no.	Gene symbol	Protein name	Localization	Disordered region (%)
8922927	C20orf46	C20orf46 protein	Membrane	21
14249536	CIRH1A	Cirhin	Nucleus	6
31874036	LYSMD3	Unnamed protein	Membrane	34
434779	RRS1	KIAA0112	Nucleus	56
2662099	KIAA0409	KIAA0409	Nucleus	40
15215317	RRP12	KIAA0690	Nucleus	25
46947023	NOL11	L14	Nucleus	13
46947031	TEX10	L18	Nucleus	14
13477303	NOL10	NOL10 protein	Nucleus	42
20806097	NOC3L	Nucleolar complex associated 3 homolog	Nucleus	42
28422560	NUP35	NUP35 protein	Nucleus	20
55664478	C1QTNF9	OTTHUMP00000018124	Cytoplasm	41
20988775	PQLC1	PQ loop repeat containing 1	Membrane	14
46255742	RBM13	RNA binding motif protein 13	Nucleus	50
13542987	(SRPS8)	Similar to ribosomal protein S8	Intracellular	53
83404927	SMU1	Smu-1 suppressor of mec-8 and unc-52 homolog	Nucleus	9
52783480	WDR43	WD-repeat protein 43	Nucleus	31
7331278	WDR12	YTM1	Nucleus	27

Proteins of which the functions are unknown but the subcellular localizations are predicted only by the electric annotation in GO were selected from Supplementary Table S1.

disordered regions [19]. The disordered region content within each amino acid sequence was calculated and compared with those of all the human proteins (25077). As can be seen in Fig. 4, the nuclear matrix fraction proteins included significantly more disordered region-rich proteins than all the human proteins ($P < 0.01$). The nuclear matrix fraction proteins could be divided into two groups based on their distribution; the two groups comprise proteins of which the disordered region contents are around 20% (Fig. 4A, left green peak, group I) and over 40% (Fig. 4A, right green peak, group II), respectively. Group I shows a similar distribution to all the human proteins. Perhaps group II is a characteristic peak of nuclear matrix proteins and comprises 110 proteins including 28 RNA binding ones (25%), 12 transcriptional factors (11%), 12 DNA-binding

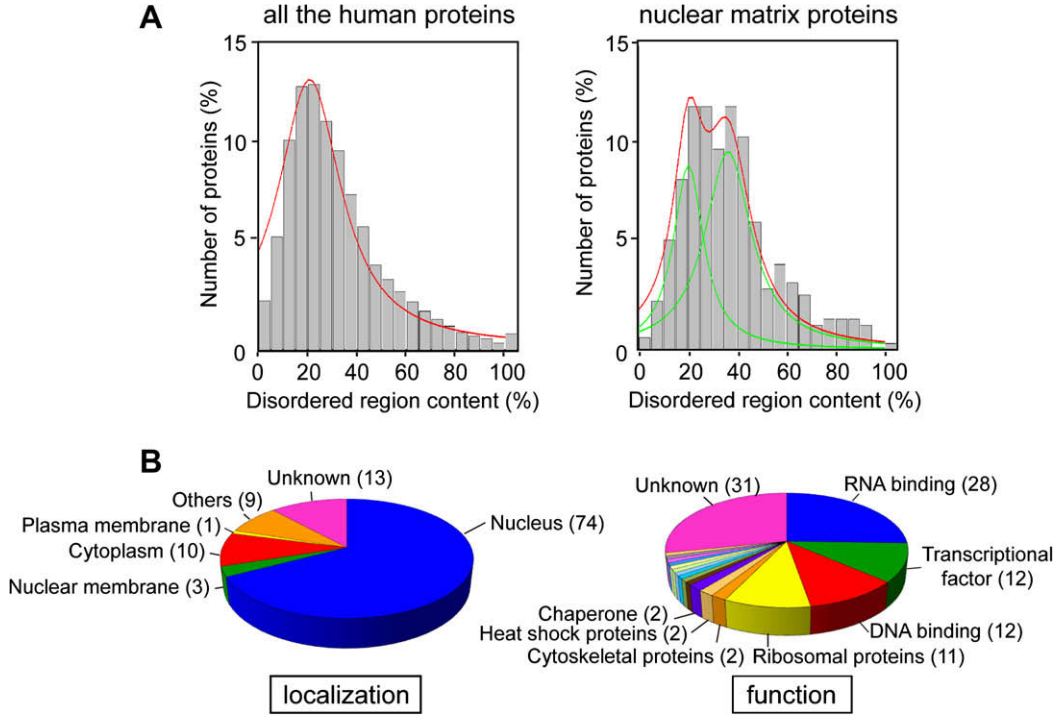

Fig. 4. Comparison of proteins present in the nuclear matrix fraction and all the human proteins as to the disordered region contents. (A) The disordered regions of 333 nuclear matrix fraction proteins (right) and all the human proteins, 25077 (left), were determined with DisEMBL, and the disordered region contents: 100 × disordered regions/(disordered regions + ordered regions), (%) were calculated. The analyzed proteins were classified according to their disordered region contents. The χ^2 values of the fitted curves are 0.964 and 0.963, respectively. (B) The localizations (left) and functions (right) of the proteins containing 40% or more disordered regions in the nuclear matrix fraction. One hundred and ten proteins containing 40% or more disordered regions in the nuclear matrix fraction were classified as to their subcellular localizations and molecular functions with Quick GO.

proteins (11%), 11 ribosomal proteins (10%), 31 unknown proteins (28%), and others (15%) (Fig. 4B). Interestingly, most of the transcriptional factors (12 of 12 proteins), DNA-binding proteins (12 of 18 proteins), and ribosomal proteins (11 of 14 proteins) identified as nuclear matrix fraction proteins are in group II (compare Fig. 4B with Fig. 1C). Thus, these proteins might function as scaffold proteins for individual functions. Group II includes 16 of the 39 novel proteins, suggesting that these 16 proteins may act as scaffold proteins in nuclear bodies and speckles or as hub molecules. Studies on the binding proteins for these proteins should provide us with a molecular basis for understanding the assembly of protein molecules in the nucleus and facilitate progress as to understanding of the inner nuclear structure.

Acknowledgements: This study was supported by a Grant-in-Aid for Scientific Research on Priority Areas (to K.T. and T.H.) from the Ministry of Education, Culture, Sports, Science and Technology of Japan. Y.H. was a recipient of a JSPS (Japan Society for the Promotion of Science) postdoctoral fellowship.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.febslet.2008.09.019.

References

[1] Stein, G.S., Zaidi, S.K., Braastad, C.D., Montecino, M., van Wijnen, A.J., Choi, J.Y., Stein, J.L., Lian, J.B. and Javed, A. (2003) Functional architecture of the nucleus: organizing the regulatory machinery for gene expression, replication and repair. Trends Cell Biol. 13, 584–592.
[2] Lamond, A.I. and Spector, D.L. (2003) Nuclear speckles: a model for nuclear organelles. Nat. Rev. Mol. Cell Biol. 4, 605–612.
[3] Cremer, T. and Cremer, C. (2001) Chromosome territories, nuclear architecture and gene regulation in mammalian cells. Nat. Rev. Genet. 2, 292–301.
[4] Pederson, T. (2000) Half a century of "the nuclear matrix". Mol. Biol. Cell 11, 799–805.
[5] Misteli, T. (2005) Concepts in nuclear architecture. Bioessays 27, 477–487.
[6] Zaidi, S.K., Young, D.W., Choi, J.Y., Pratap, J., Javed, A., Montecino, M., Stein, J.L., van Wijnen, A.J., Lian, J.B. and Stein, G.S. (2005) The dynamic organization of gene-regulatory machinery in nuclear microenvironments. EMBO Rep. 6, 128–133.
[7] Fey, E.G., Krochmalnic, G. and Penman, S. (1986) The nonchromatin substructures of the nucleus: the ribonucleoprotein (RNP)-containing and RNP-depleted matrices analyzed by sequential fractionation and resinless section electron microscopy. J. Cell Biol. 102, 1654–1665.
[8] Segawa, M., Niino, K., Mineki, R., Kaga, N., Murayama, K., Sugimoto, K., Watanabe, Y., Furukawa, K. and Horigome, T. (2005) Proteome analysis of a rat liver nuclear insoluble protein fraction and localization of a novel protein, ISP36, to compartments in the interchromatin space. FEBS J. 272, 4327–4338.
[9] Yoshida, Y., Miyazaki, K., Kamiie, J., Sato, M., Okuizumi, S., Kenmochi, A., Kamijo, K., Nabetani, T., Tsugita, A., Xu, B., Zhang, Y., Yaoita, E., Osawa, T. and Yamamoto, T. (2005) Two-dimensional electrophoretic profiling of normal human kidney glomerulus proteome and construction of an extensible markup language (XML)-based database. Proteomics 5, 1083–1096.
[10] Haze, K., Yoshida, H., Yanagi, H., Yura, T. and Mori, K. (1999) Mammalian transcription factor ATF6 is synthesized as a transmembrane protein and activated by proteolysis in response to endoplasmic reticulum stress. Mol. Biol. Cell 10, 3787–3799.
[11] Enari, M., Ohmori, K., Kitabayashi, I. and Taya, Y. (2006) Requirement of clathrin heavy chain for p53-mediated transcription. Genes Dev. 20, 1087–1099.
[12] Li, D. and Roberts, R. (2001) WD-repeat proteins: structure characteristics, biological function, and their involvement in human diseases. Cell Mol. Life Sci. 58, 2085–2097.
[13] Venter, J.C. et al. (2001) The sequence of the human genome. Science 291, 1304–1351.
[14] Chang, B.Y., Chiang, M. and Cartwright, C.A. (2001) The interaction of Src and RACK1 is enhanced by activation of protein kinase C and tyrosine phosphorylation of RACK1. J. Biol. Chem. 276, 20346–20356.
[15] Garcia-Higuera, I., Fenoglio, J., Li, Y., Lewis, C., Panchenko, M.P., Reiner, O., Smith, T.F. and Neer, E.J. (1996) Folding of proteins with WD-repeats: comparison of six members of the WD-repeat superfamily to the G protein beta subunit. Biochemistry 35, 13985–13994.
[16] Smith, T.F., Gaitatzes, C., Saxena, K. and Neer, E.J. (1999) The WD repeat: a common architecture for diverse functions. Trends Biochem. Sci. 24, 181–185.
[17] Hennig, L., Bouveret, R. and Gruissem, W. (2005) MSI1-like proteins: an escort service for chromatin assembly and remodeling complexes. Trends Cell Biol. 15, 295–302.
[18] Dragon, F., Gallagher, J.E.G., Compagnone-Post, P.A., Mitchell, B.M., Porwancher, K.A., Wehner, K.A., Wormsley, S., Settlage, R.E., Shabanowitz, J., Osheim, Y., Beyer, A.L., Hunt, D.F. and Baserga, S.J. (2002) A large nucleolar U3 ribonucleoprotein required for 18S ribosomal RNA biogenesis. Nature 417, 967–970.
[19] Tompa, P. (2002) Intrinsically unstructured proteins. Trends Biochem. Sci. 27, 527–533.
[20] Dunker, A.K., Cortese, M.S., Romero, P., Iakoucheva, L.M. and Uversky, V.N. (2005) Flexible nets. The roles of intrinsic disorder in protein interaction networks. FEBS J. 272, 5129–5148.
[21] Nakamura, T., Mori, T., Tada, S., Krajewski, W., Rozovskaia, T., Wassell, R., Dubois, G., Mazo, A., Croce, C.M. and Canaani, E. (2002) ALL-1 is a histone methyltransferase that assembles a supercomplex of proteins involved in transcriptional regulation. Mol. Cell 10, 1119–1128.
[22] Wang, Y., Cortez, D., Yazdi, P., Neff, N., Elledge, S.J. and Qin, J. (2000) BASC, a super complex of BRCA1-associated proteins involved in the recognition and repair of aberrant DNA structures. Genes Dev. 14, 927–939.